NASA Heliophysics Data Environment (HPDE): Recent Progress

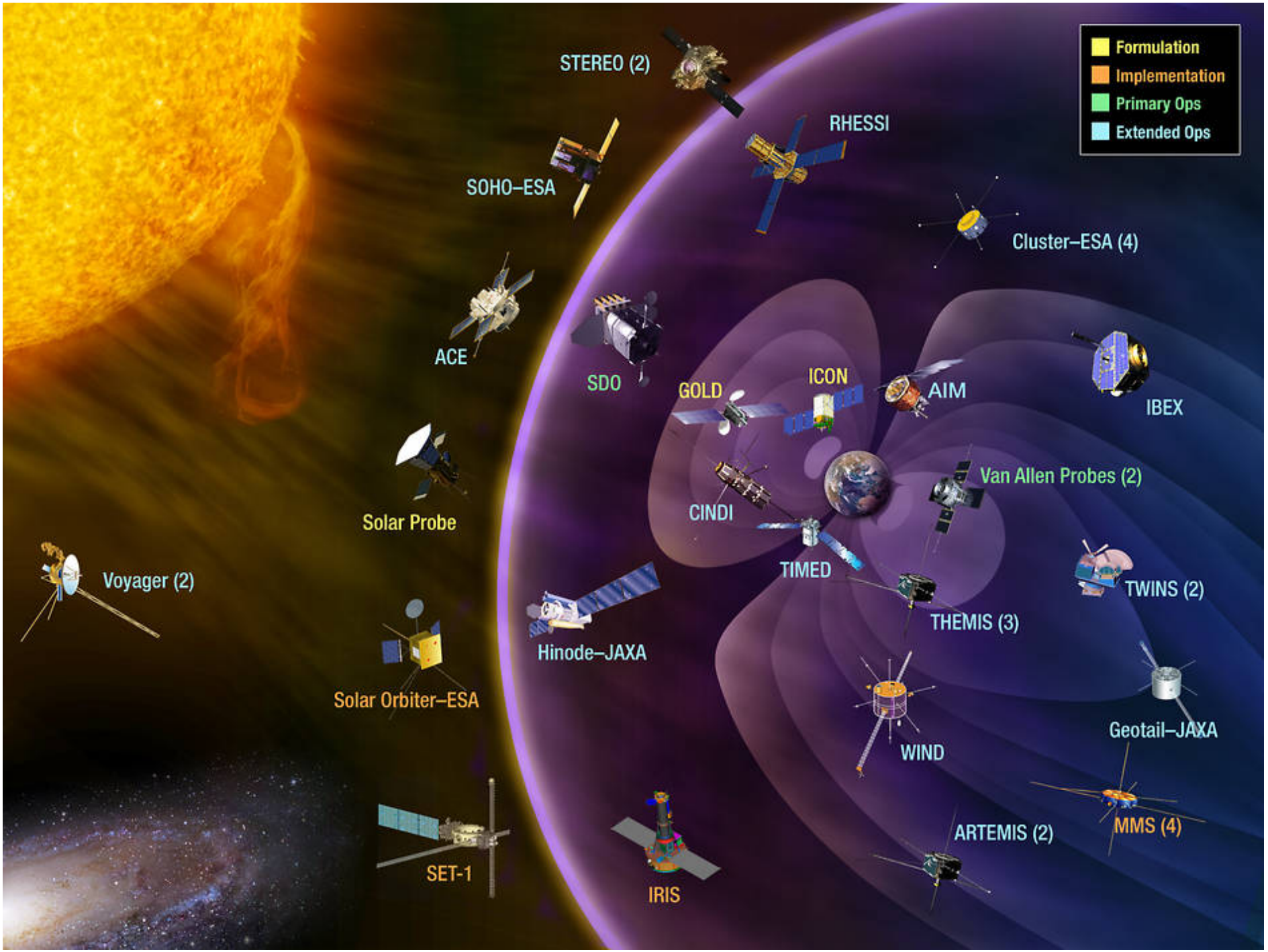
D. Aaron Roberts

NASA GSFC

HDMC Project Scientist

HPDE Program Scientist

CCMC April 2018 Workshop


Formulation
Implementation
Primary Ops
Extended Ops
STEREO (2)
RHESSI
SOHO–ESA
Cluster–ESA (4)
ACE
SDO
GOLD
ICON
AIM
IBEX
Van Allen Probes (2)
CINDI
Solar Probe
TIMED
TWINS (2)
Voyager (2)
THEMIS (3)
Hinode–JAXA
Solar Orbiter–ESA
Geotail–JAXA
WIND
MMS (4)
ARTEMIS (2)
SET-1
IRIS

Overview

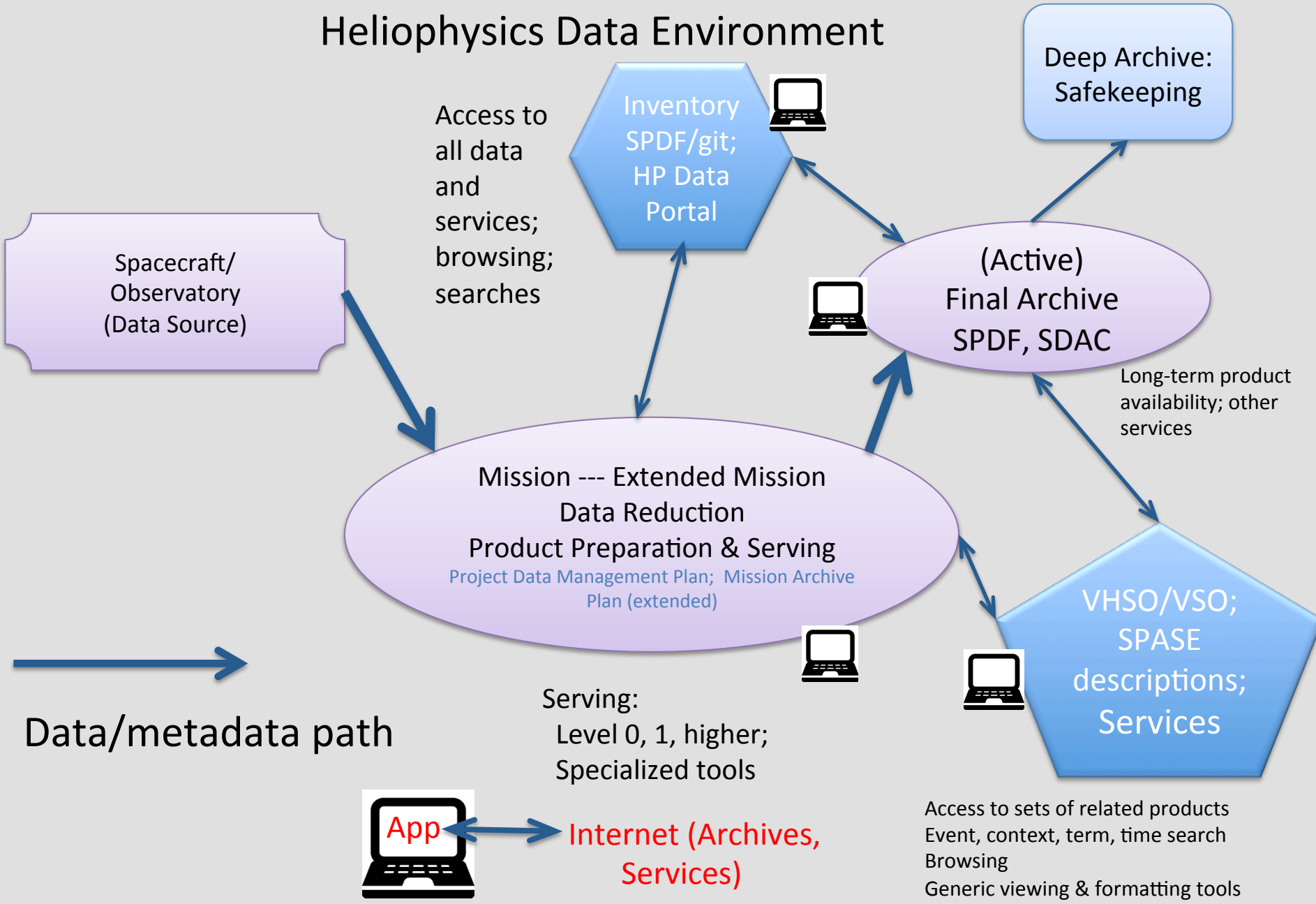
- *Really Brief history of the HP Data Environment (HPDE)*
- *Current Status*: Archives, Registry, Web (browser, ftp) and Web Service Access, Tools
- *Role of standards*
- *Current standards-based tools*
- *Remaining challenges*

See http://hpde.gsfc.nasa.gov (includes the Data Policy) and http://heliophsyicsdata.gsfc.nasa.gov (generic data access via the "Heliophysics Data Portal")

History

- *Brief history of the HP Data Environment*:
  - From "lab benches in space" to an *open*, comprehensive, searchable, accessible, distributed archive.
  - From "give something to NSSDC" as an archiving plan to a coherent NASA HP Science Data Management Policy (orig. 2007; *v1.1* 2009: Archives codified; *v 1.2* 2016: consolidate "VxOs").
  - Current distributed archives: discover and access data from hundreds of observatories (spacecraft, etc.) covering more than 50 years of exploration; many datasets are at SPDF or SDAC (active final archives).
  - Current services and development support the "DRIVE" initiative of the NRC Decadal Survey.

Heliophysics Data Environment



Value Added Services use this framework

Virtual Observatory definitions (after "VESPA," a European VO)

- "A virtual observatory": a web-based *portal* providing access to remotely distributed data resources; *vs.*
- *"The virtual observatory": **standards** to allow easy, uniform access to data [and models] and to enable **interoperable tools** to share data transparently. "The VO" is **the invisible machinery that allows us to work efficiently.***

STANDARDS, STANDARDS, STANDARDS

- Data format (FITS, CDF, NetCDF, [HDF]): Self-describing, with standard metadata for easy reading by standard tools.
- Basic registration metadata terms and format (SPASE): Easy access to all datasets; inventory tracking.
- General API for data access (HAPI): Let all tools have easy access to data, independent of format.

GODDARD SPACE FLIGHT CENTER
Space Physics Data Facility

+ Goddard Home
+ Visit NASA.gov

A VO Face of the Inventory
Uses standard metadata

Heliophysics Data Portal "Find it. Browse it. Get it."

SPASE inside

Help | Geo Orbits | Helio Orbits | SPASE Registry | ADS Abstracts | Weather | Feedback

Text Restriction

Add

Time Span Restriction

YYYY-MM-dd or YYYY-DDD

from:

to: Add

Element Restriction

- Resource type
- Measurement type
- Observatory Group
- Observatory (~200!)
- Instrument
- Observed region
- Spectral range
- Cadence
- Repository Name
- Access rights
- Format

Current Product Restrictions

No restrictions are currently set.

Showing 1 - 20 of 1601 Results | View Current List | Sort by Observatory

#	Products (& SPASE descriptions)	Access Links
1	ACE 27-day Survey Plots	· Polar-Wind-Geotail 'gif-walk' site Get Images/Plots
2	ACE Cosmic Ray Isotope Spectrometer (CRIS) 1-Hour Level 2 Data	· FTP access to files at SPDF · HTTP access to files at SPDF · CDAWeb Get Data/Plots · ACE Science Center FTP area · ACE Science Center · ACE Cosmic Ray Isotope Spectrometer (CRIS) Level 2 data Home Page
3	ACE CRIS L2 1-day Z=3-28 flux data	· ACE Science Center · ACE/CRIS L2 data in HDF via ftp · CDAWeb Get Data/Plots · FTP access to files at SPDF · HTTP access to files at SPDF
4	ACE Daily Survey Plots	· Polar-Wind-Geotail 'gif-walk' site Get Images/Plots
5	ACE Electron Proton Alpha Monitor (EPAM) 1-hour Key Parameter data	· FTP access to files at SPDF · HTTP access to files at SPDF · CDAWeb Get Data/Plots

Directly Read Data from CDAWeb into IDL (Common approach; HAPI generalizes this)

```
timename='jul_day'    ;name of time variable -- Julian days
start_time = '1998-06-10T00:00:00.0Z'  ;start time
stop_time  = '1998-06-10T23:59:59.0Z'  ;stop time
dt_sec=10.0 ;sec -- bin size in seconds

dataset_id='WI_H0_MFI'    ; CDAWeb dataset ID
vars=[ 'B3F1=Bmag3', 'B3GSE=Bx3,By3,Bz3']   ; CDAWeb variable names with locally assigned names
```

cdaweb_get_bin, dataset_id,vars,start_time, stop_time,dt_sec,time_name=timename

```
dataset_id='WI_PM_3DP'
vars=[ 'P_DENS=np3', 'P_VELS=Vxp3,Vyp3,Vzp3',  'P_TEMP=Tp3', 'A_DENS=na3',  $
          'A_VELS=Vxa3,Vya3,Vza3', 'A_TEMP=Ta3']
```

cdaweb_get_bin, dataset_id,vars,start_time, stop_time,dt_sec,time_name=timename,/autobad

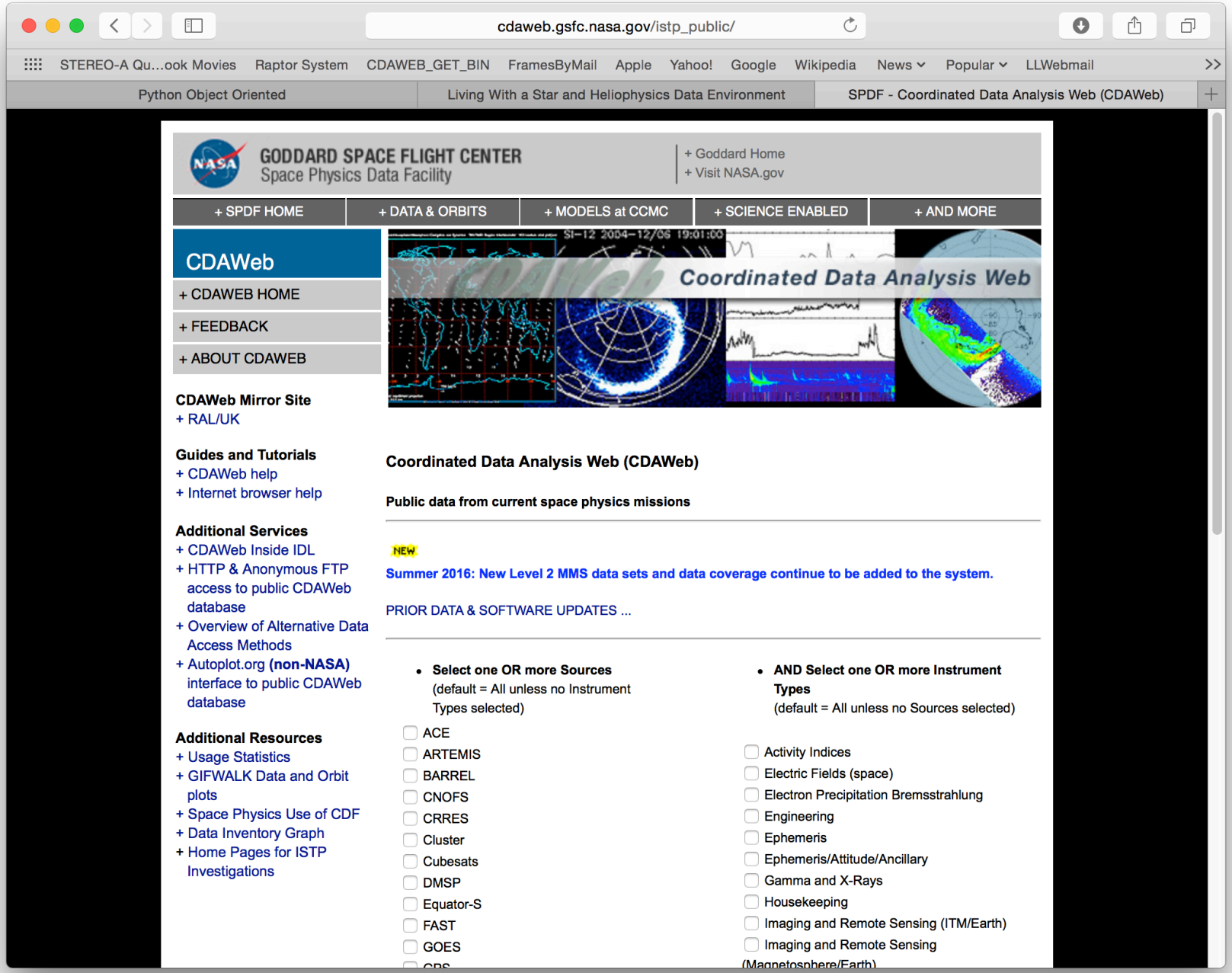
No more writing code for every dataset.

The Internet functions as a local, easy to use hard drive.

Prototype for more general “Virtual Observatory” access to any HP data.

VSO does the same for Solar Data; “AMDA” (Europe) uses SPDF web services

cdaweb.gsfc.nasa.gov



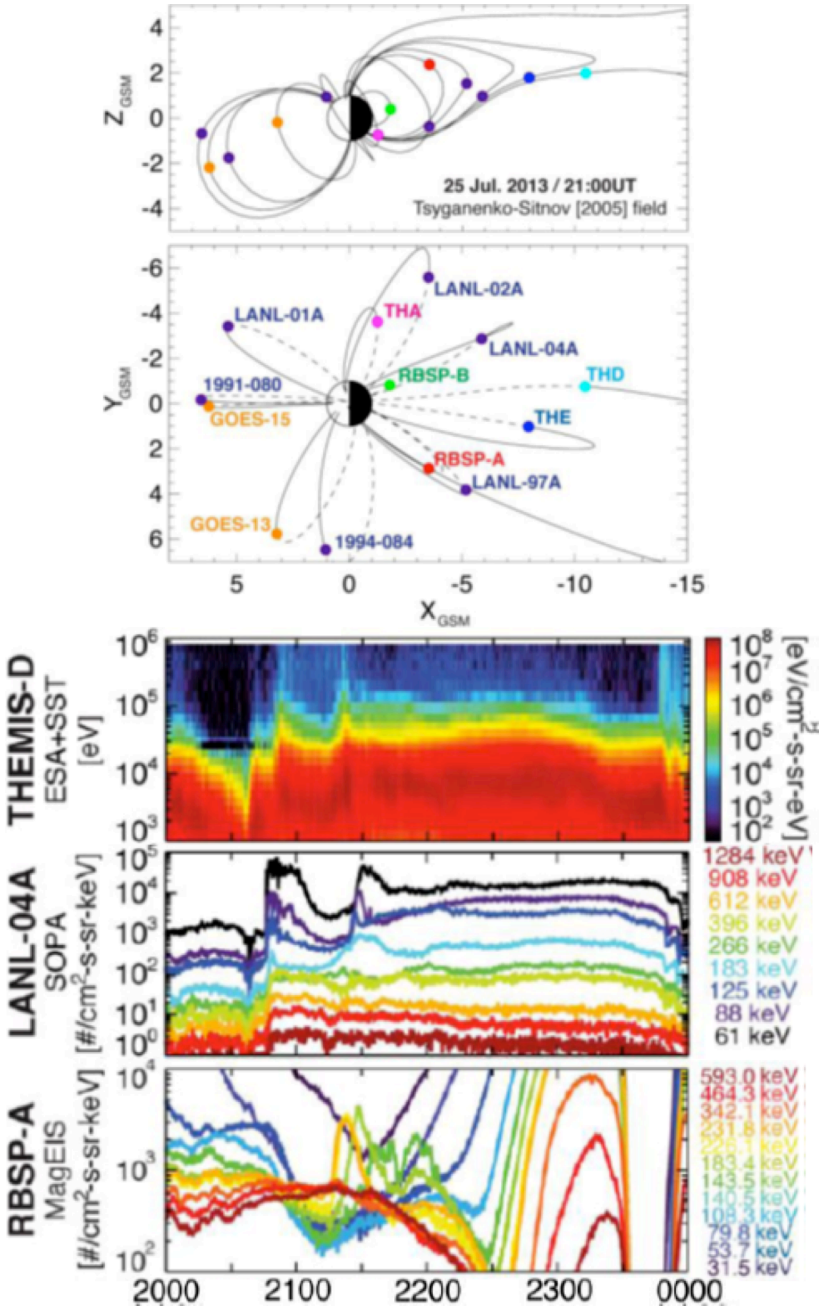
SPEDAS (Space Physics Environment Data Analysis Software)

spedas.org

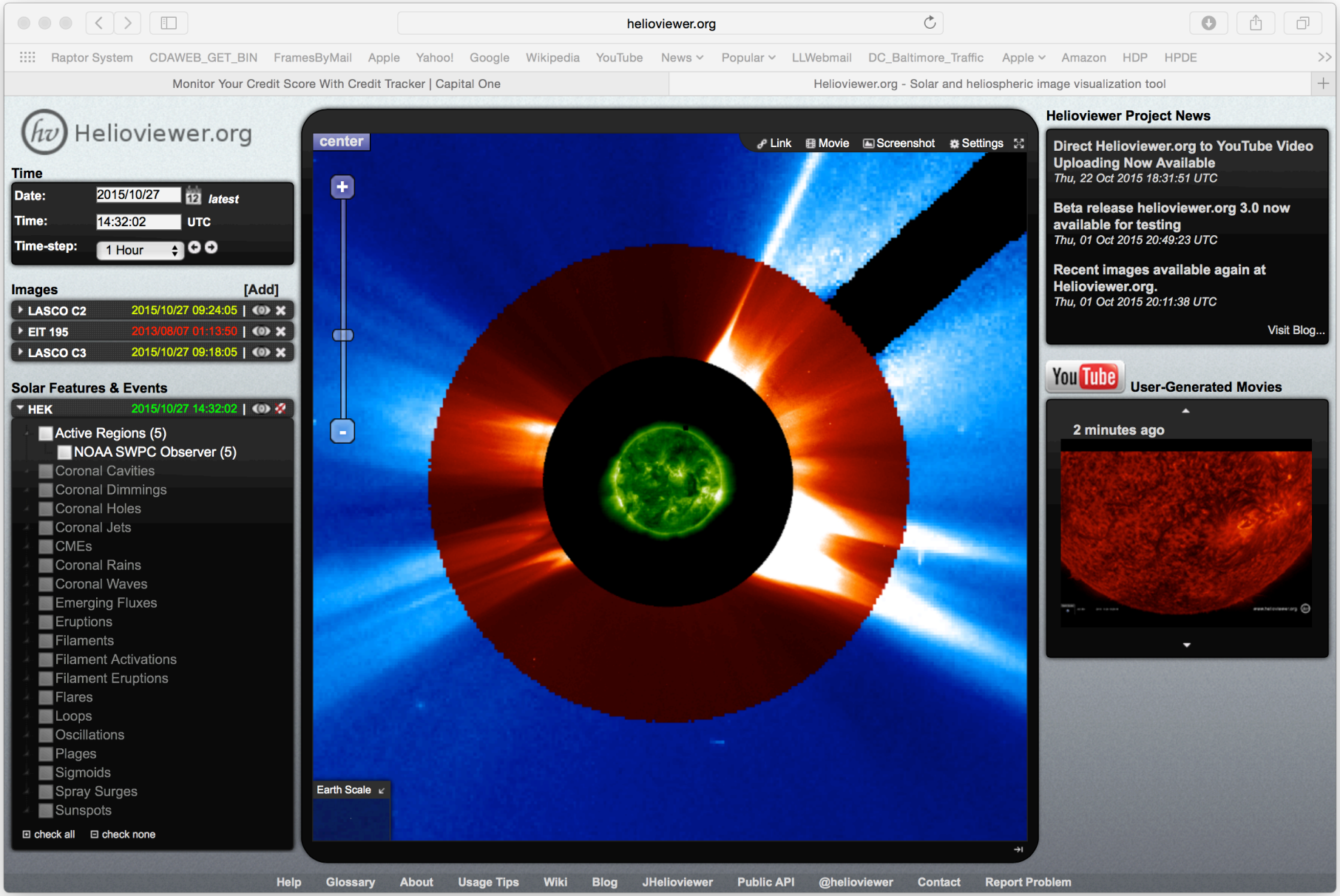
Generic IDL analysis tool with a stand-alone version and with direct links to much of Space Physics data, including all of SPDF.

Roughly the Space Physics equivalent of SolarSoft (supported via SDAC).

Developed by various missions, most recently THEMIS. Used now by MMS and many others.



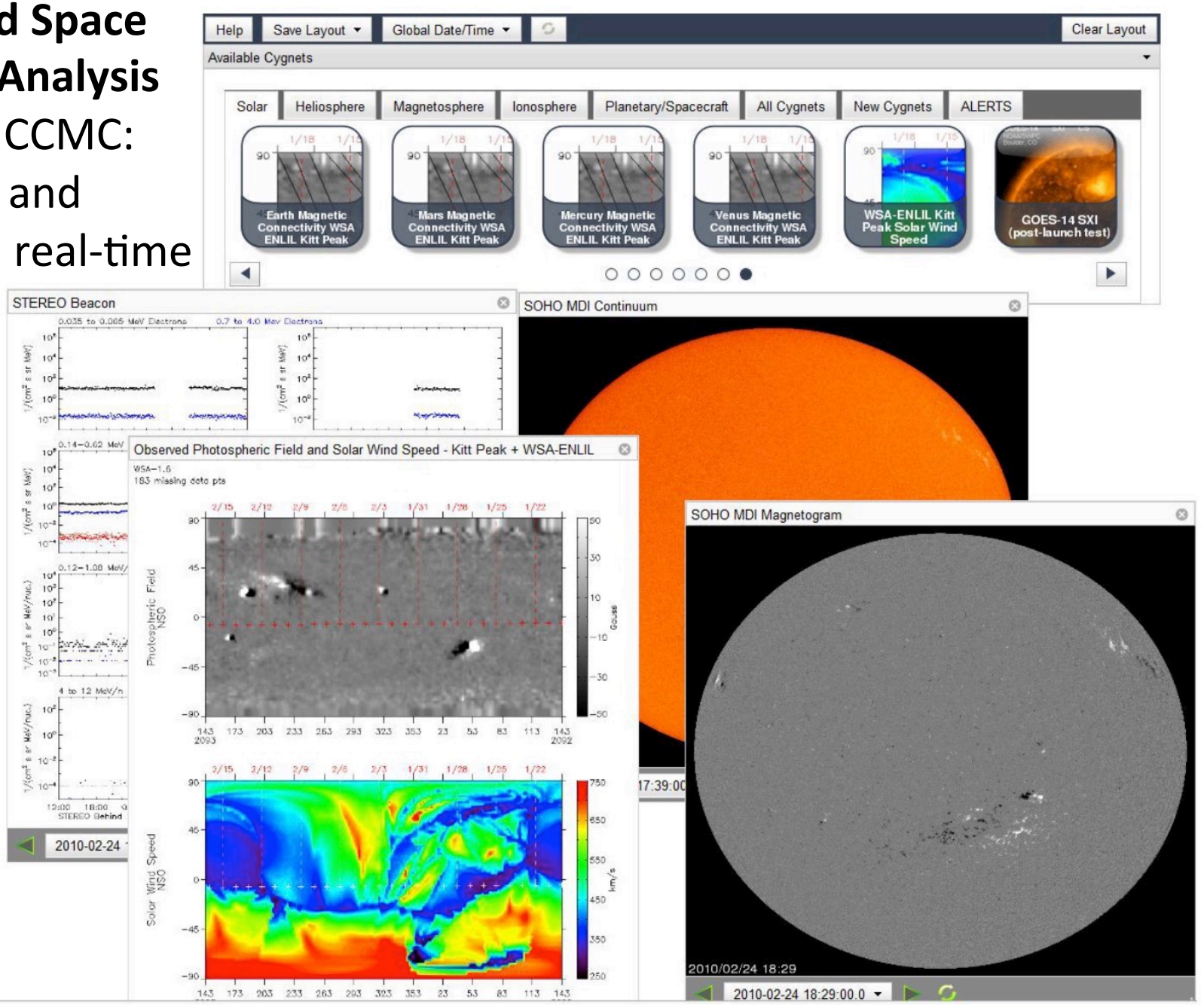
Helioviewer: Multispacecraft solar imaging and movie-making



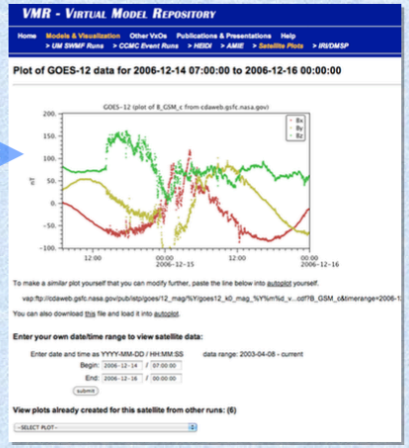
Autoplot (autoplot.org): Reads multiple formats including [Complex ASCII tables; Binary tables; Common Data Format (CDF); NcML; SPASE; Cluster Exchange Format; NetCDF; OpenDAP; HDF5; TSDS; FITS; Excel; Wav; PNG, JPG, etc. For details and a full list, see Formats.]

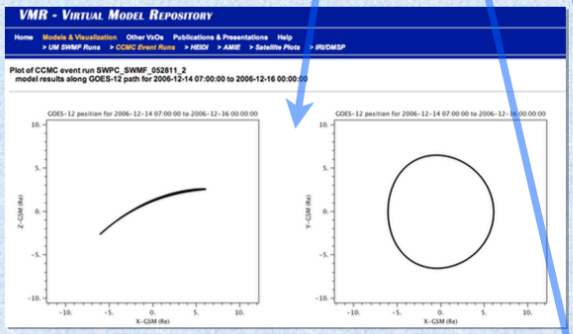
http://iswa.gsfc.nasa.gov

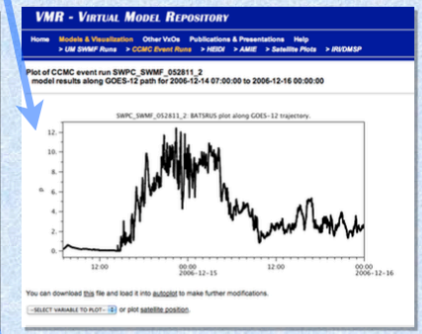
Integrated Space Weather Analysis (iSWA) at CCMC: Observed and simulated real-time data

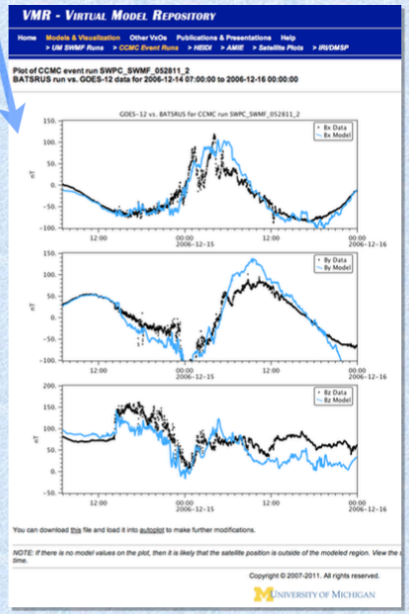


Multiple plot types are available:
- satellite B field
- model output along satellite trajectory
- data/model comparison
- satellite location









The data/model plot on the right includes many parts:
- model run (CCMC)
- satellite trajectory (SSCWeb)
- model interpolation (CCMC)
- satellite data (CDAWeb)
- collect plot pieces (VMR)
- plotting (Autoplot)

http://vmr.engin.umich.edu
Now CCMC

Data-Model Comparisons

Current HDMC effort: "Virtual Model Repository"

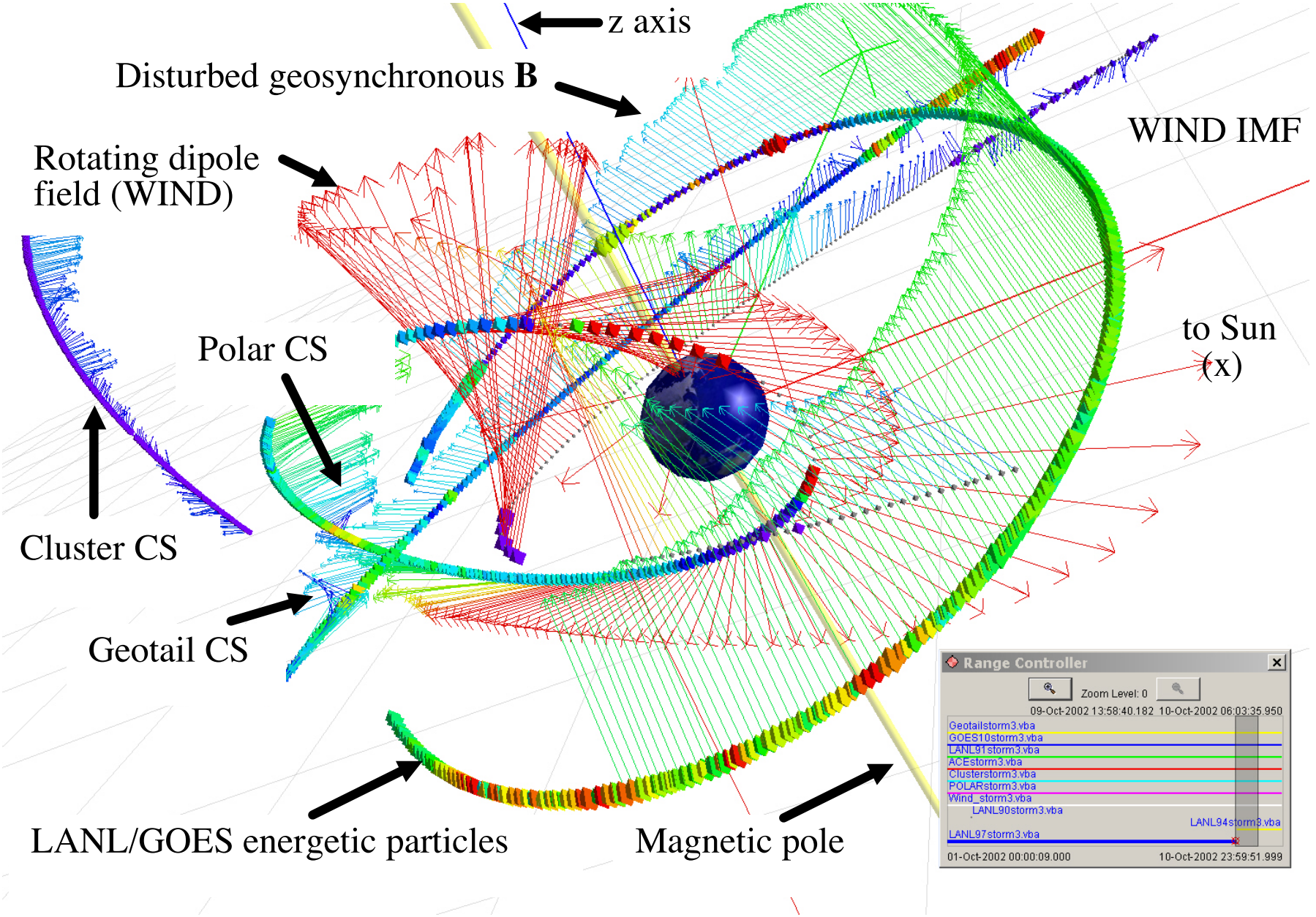
Uses Internet connectivity to load spacecraft data and simulations into the same visualization.

A way forward:

Develop a community-based plan for generalizing this work.

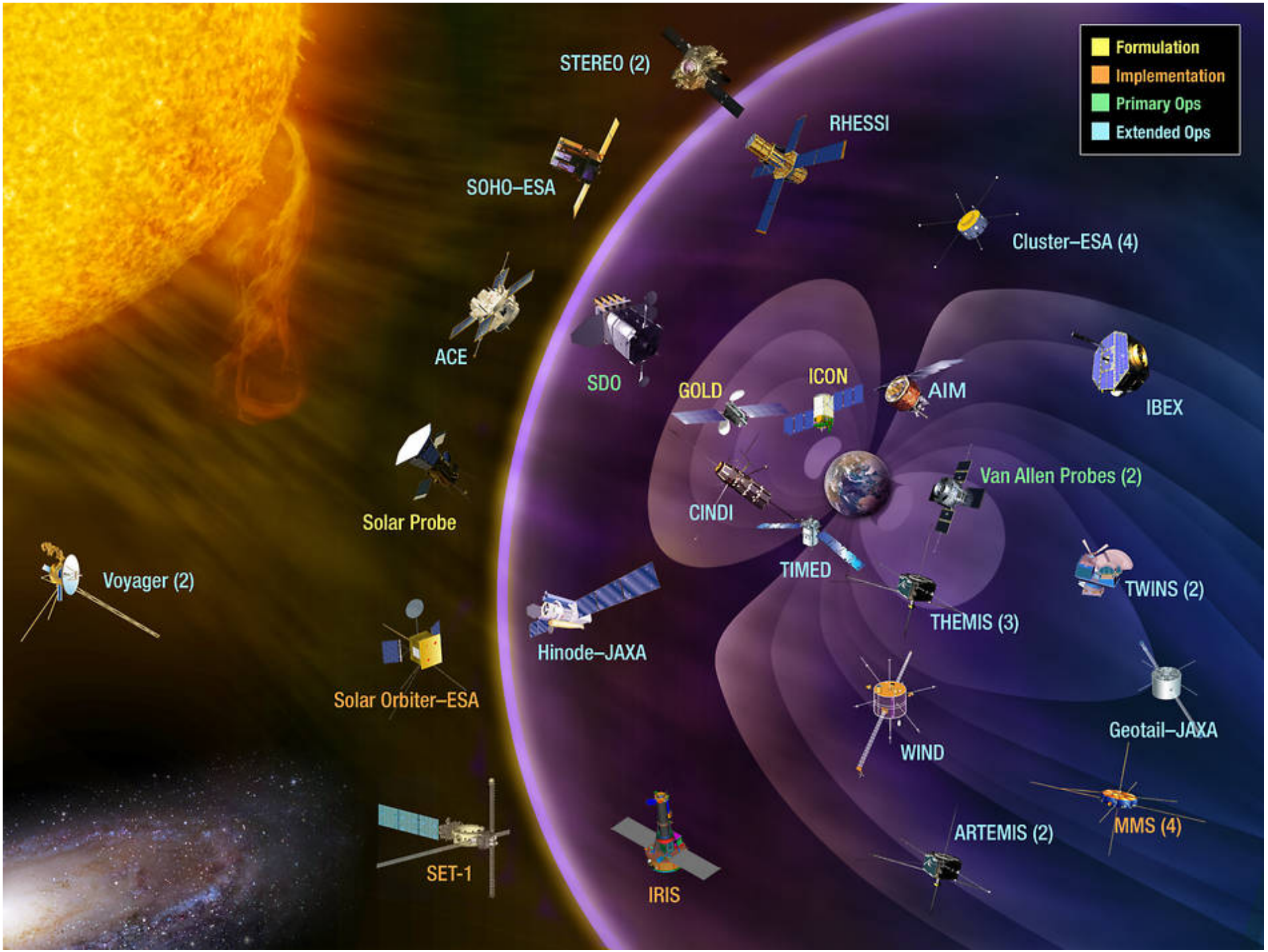
Start with the CCMC to develop ideas and come up with a cohesive approach.

Fund specific efforts to implement the plan.


z axis
Disturbed geosynchronous **B**
Rotating dipole field (WIND)
WIND IMF
to Sun (x)
Polar CS
Cluster CS
Geotail CS
LANL/GOES energetic particles
Magnetic pole
Range Controller
Zoom Level: 0
09-Oct-2002 13:58:40.182
10-Oct-2002 06:03:35.950
Geotailstorm3.vba
GOES10storm3.vba
LANL91storm3.vba
ACEstorm3.vba
Clusterstorm3.vba
POLARstorm3.vba
Wind_storm3.vba
LANL90storm3.vba
LANL94storm3.vba
LANL97storm3.vba
01-Oct-2002 00:00:09.000
10-Oct-2002 23:59:51.999

Some Major Future Challenges

- **Large data volumes**
  - How to use the data: *Pattern recognition; data mining; 4-D visualization*
- **Model-data comparisons and insights**
  - *Seamless integration of model output with data streams*
  - Data assimilation; *true space weather capabilities*


Formulation
Implementation
Primary Ops
Extended Ops
STEREO (2)
RHESSI
SOHO–ESA
Cluster–ESA (4)
ACE
SDO
GOLD
ICON
AIM
IBEX
Van Allen Probes (2)
CINDI
Solar Probe
TIMED
TWINS (2)
Voyager (2)
THEMIS (3)
Hinode–JAXA
Solar Orbiter–ESA
Geotail–JAXA
WIND
MMS (4)
ARTEMIS (2)
SET-1
IRIS

Additional Material

- Slides from the HPDE presentation to the UMCP Space Weather Workshop, Fall, 2016 and the CCMC Workshop, Spring, 2016

Overview

- *Brief history of the HP Data Environment*:
  - From "lab benches in space" to an *open*, comprehensive, searchable, accessible, distributed archive;
  - From "give something to NSSDC" to a coherent NASA HP Science Data Management Policy (orig. 2007; *v1.1* 2009: Archives codified; *v 1.2* 2016: consolidate "VxOs")
- *Current Status*: Archives, Registry, Web (browser, ftp) and Web Service Access
- *Plans*: (Active) Final Archives and the HDMC
  - Focus on (international) standards for data formats, metadata formats, and a data API to allow uniform access to both data and models from a variety of archives.
  - Develop successively more powerful tools based on the standards.
- See http://hpde.gsfc.nasa.gov (includes the Data Policy) and http://heliophsyicsdata.gsfc.nasa.gov (generic data access via the "Heliophysicis Data Portal")

STANDARDS: Generation and Maintenance

- Data format (FITS, CDF, NetCDF): SPDF, HEASARC, solar community, TIMED/ICON/GOLD with SPDF
- Basic registration metadata terms and format (SPASE): Upkeep and generation for new and legacy missions (core group exists to do this)
- API for data access (in progress): Use SPDF, VSO, Autoplot, and others as guides (these work now); syntax and some instances are in progress.
- Standard tools: e.g. Autoplot, SPEDAS, Helioviewer

Heliophysics Data Environment Current Status

• The VxOs and the HP Data Centers have, over recent years, *produced a comprehensive set of dataset descriptions* in standard terms, thus allowing *ready access to most Heliophysics data* (e.g., through SPDF's HP Data Portal). The *Virtual Observatory* paradigm is "universal access through standards."

• HP Final Archives (SDAC, SPDF) have been systematically *working with new missions* (RBSP, BARREL, MMS, Solar Probe, IRIS) to assure that *data will be easily available for the short and long term.*

• *Web (SPDF, VSO, etc.) services facilitate direct access* to most HP data from applications such as IDL, and from others' services (used by EU AMDA; data/model services; SPEDAS).

• We have *established **standards** for data formats* (FITS, CDF, and NetCDF), and for *metadata* (SPASE), and these are stable but responsive to community needs.

• We have *restored a large fraction of the data from older missions* (ISEE-1, 2, 3; DE 1, 2; Helios 1, 2; AMPTE; FAST; Yohkoh; etc.), and these data are accessible in standard formats from the HP Archives. We are systematically examining the *NSSDC data archives* and restoring datasets of scientific interest.

• *Data are no longer dropped when missions end*, and even the "Resident Archives" that were needed for this are becoming nearly obsolete.

The Overall Plan

- Phase out the current "VxOs" in favor of a unified "VHSO" based on standards for data and metadata formats and for a data API.
- Have one central group in charge of the metadata standards (content, format, and utilities) and related provision of data product documentation.
- Have one group in charge of a data API and a reference VHSO implementation. ("VHSO" = Virtual HSO.)
- Add selected Value Added Services (e.g, visualization, data mining) as they become mature (a few are now; see below).
- Reinstitute a Value Added Services line in the HDEE (now data upgrades) call to allow for innovation.

The Data Policy states various functions for the HP Data Environment

- Produce and serve high-quality, well-documented data

- Provide open access to scientifically useful data products
  - Allow easy discovery of all available products and their location
  - Provide easily useable, well-documented products
  - Provide uniformity of access to data

- Keep data flowing without interruption when missions end
  - Provide funds to continue post-mission serving of data, if needed
  - Move data to Active Final Archives for long-term serving

- Keep data safe for the long term
  - Assure data are safe at all stages
  - Provide long-term archives for safe-keeping

A "Virtual Observatory" approach to sharing of model and simulation output, and to data-model comparisons: Outlook on CCMC support of VOs

Aaron Roberts, NASA GSFC
(Heliophysics Data and Model Consortium)
Jeffrey Hayes, NASA HQ
Darren De Zeeuw, U Michigan
(Virtual Model Observatory)

The Problem

- Many researchers want to be able to access and use data from simulations and models, often in conjunction with spacecraft data.
- Many modelers want to be able to easily share their results with others.
- The above are being done now on an *ad hoc* basis (outside of CCMC); uniformity of access and use would make everyone's research more efficient.
- CCMC is working toward such standards, but we need buy-in and a general plan.
- Additional resources are available, but we need a plan to make them effective.

NASA and More General Context

NRC Decadal Survey: DRIVE initiative

- *A relatively small, low-cost initiative, **DRIVE** provides high leverage to current and future space science research investments with a diverse set of science-enabling capabilities.*
- **D**iversify observing platforms with microsatellites and midscale ground-based assets.
- **Realize scientific potential by sufficiently funding operations and data analysis.**
- **Integrate observing platforms and strengthen ties between agency disciplines**.
- **V**enture forward with science centers and instrument and technology development.
- **E**ducate, empower, and inspire the next generation of space researchers.

Decadal Survey: Overview

- Significant progress has been made over the last decade in establishing the essential components of the solar and space physics data environment. However, to achieve key national research and applications goals*, a data environment that draws together new and archived satellite and ground-based solar and space physics data sets and computational results from the research and operations communities is needed*.

Decadal Survey: Future Goals

- Heliophysics is poised to make a natural transition from being driven predominantly by the pursuit of basic scientific understanding of physical processes towards one that must also address more operational, application-specific needs, much like terrestrial weather forecasting. *This transition requires **(1) instant unfettered access to a wide array of datasets from distributed sources in a uniform, standardized format, (2) incorporation of the results of community-developed models,** and (3) the ability to perform simulations interactively and to couple different models to track ongoing space-weather events.*
- NASA has already taken the important first step in integrating many of these datasets and tools to form the Heliophysics Data Environment (HPDE). The main objective of the HPDE is to implement a distributed, integrated, flexible data environment. HPDE modeling centers should serve as a sound foundation for a future, ***fully integrated heliophysics data and modeling center***.

VO Required Standards

- Data formats: e.g., "CDF" with specific conventions ("ISTP" "Kameleon")
- Data product description (metadata) standards, e.g. the "SPASE" ("IMPEx" version for simulations); XML implementation
- Data transfer protocols (APIs, streaming format)
- ***All the above must be implemented (all relevant objects described) or this is useless; both agreement on standards and implementation are difficult issues***.
- (A related issue is how to deal with "big data" and server-side tools.)

How to proceed?

- A substantial part of the community should *agree on the goal*
- Standards need to be *developed/adopted*
  - CCMC has developed and implemented the Kameleon standard for output
  - CCMC is implementing SPASE/IMPEx as a metadata standard
- Community *buy-in* and implementation can lead to new *tool development* and application
- Similar standards for spacecraft data allow *data-model integration* (e.g., CCMC and "VMR" from U Michigan).